2018

Combinatorial algorithms and methods for security of statistical databases related to the work of Mirka Miller

A V. Kelarev
University of Newcastle

Jennifer Seberry
University of Wollongong, jennie@uow.edu.au

Leanne Rylands
Universty of Western Sydney

Xun Yi
Royal Melbourne Institute of Technology

Combinatorial algorithms and methods for security of statistical databases related to the work of Mirka Miller

Abstract

This article gives a survey of combinatorial algorithms and methods for database security related to the work of Mirka Miller. The main contributions of Mirka Miller and coauthors to the security of statistical databases include the introduction of Static Audit Expert and theorems determining time complexity of its combinatorial algorithms, a polynomial time algorithm for deciding whether the maximum possible usability can be achieved in statistical database with a special class of answerable statistics, NP-completeness of similar problems concerning several other types of databases, sharp upper bounds on the number of compromise-free queries in certain categories of statistical databases, and analogous results on applications of Static Audit Expert for the prevention of relative compromise.

Disciplines

Engineering | Science and Technology Studies

Publication Details

Kelarev, A., Seberry, J., Rylands, L. & Yi, X. (2018). Combinatorial algorithms and methods for security of statistical databases related to the work of Mirka Miller. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10765 LNCS 383-394. International Workshop on Combinatorial Algorithms (IWOCA 2017)

Combinatorial algorithms and methods for security of statistical databases related to the work of Mirka Miller

Andrei Kelarev[1], Jennifer Seberry[2], Leanne Rylands[3], Xun Yi[1]

[1]School of Science, RMIT University
GPO Box 2476, Melbourne, VIC 3001, Australia
Email: andrei.kelarev@gmail.com,xun.yi@rmit.edu.au

[2]School of Computing and Information Technology
University of Wollongong, Northfields Avenue, NSW 2522, Australia
Email: jennie@uow.edu.au

[3]School of Computing, Engineering and Mathematics
Western Sydney University, Locked Bay 1797, Penrith, NSW 2751, Australia
Email: l.rylands@westernsydney.edu.au

Abstract. This article gives a survey of combinatorial algorithms and methods for database security related to the work of Mirka Miller. The main contributions of Mirka Miller and coauthors to the security of statistical databases include the introduction of Static Audit Expert and theorems determining time complexity of its combinatorial algorithms, a polynomial time algorithm for deciding whether the maximum possible usability can be achieved in statistical database with a special class of answerable statistics, NP-completeness of similar problems concerning several other types of databases, sharp upper bounds on the number of compromise-free queries in certain categories of statistical databases, and analogous results on applications of Static Audit Expert for the prevention of relative compromise.

Keywords: combinatorial algorithms, NP-completeness, privacy in data mining, database security, time complexity, sharp upper bounds.

1 Introduction

This article surveys combinatorial algorithms and methods for maintaining the security of statistical databases. We include concise statements of the main theorems and results related to the work of Mirka Miller. For background information and preliminaries, the readers are referred to [6], [17], [18], [42] and [48]. An excellent overview of various concepts used in statistical disclosure control with detailed explanations and examples illustrating the major notions is given in [4].

Statistical databases are databases in which only statistical types of queries are allowed. They store records with data on individuals (companies, organizations, etc) and can output statistics concerning subsets of individuals providing

aggregated information on groups of records in the database, while protecting confidential data of individuals from disclosure. Users can pose statistical queries, which are either answered (precisely or approximately), or rejected by a control mechanism to ensure the privacy of confidential data of individuals. Statistical databases are very important for numerous practical application. For example, answers to statistical queries can help medical researchers to evaluate the effectiveness of medications or certain lifestyle changes for the treatment or prevention of various conditions.

It is usually possible to deduce confidential information by comparing the results of several different queries. The security problem for statistical databases is to develop control mechanisms that will prevent direct or indirect disclosure of confidential data by the release of statistics as answers to statistical queries.

2 Classical compromise

If the value of a protected attribute of an individual record can be derived, then the database is said to have been (positively) *compromised*. It is shown in [33] how supplementary knowledge available from other sources can be exploited to obtain values of a confidential attribute. The following types of supplementary knowledge are defined in [33]. *Supplementary knowledge of type I* is knowledge of the values of attributes which uniquely identify a particular record or a particular subset of records in a database. *Supplementary knowledge of type II* is knowledge of the value of a confidential attribute for a particular individual.

Let us denote the numerical attributes contained in a statistical database by A_0, $A_1, \ldots$, A_m. Without loss of generality we may assume that the users can submit queries on statistics concerning the attribute A_0 and the values of attributes $A_1, \ldots, A_m$ are used to select subsets of records for these queries. Then A_0 is called a *quantitative attribute* and $A_1, \ldots, A_m$ are called *characteristic attributes* for such queries. The set of records chosen for a query by specifying conditions on the characteristic attributes is called the *query set*. Denote by n the number of records stored in the database. Let x_1, $x_2, \ldots$, x_n be the (protected) values of the quantitative attribute in these records.

A *SUM query* is a sum of the form $a_1x_1 + \cdots + a_nx_n$, where $a_i = 1$ if the i-th record belongs to the query set, and $a_i = 0$ otherwise. For SUM queries, it is enough to consider 1-dimensional statistical databases, or databases with only one quantitative attribute. An arbitrary set of SUM queries in a multi-dimensional statistical database can be represented as a disjoint union of SUM queries corresponding to different quantitative attributes, and each of these subsets can be viewed as a set of SUM queries of the corresponding 1-dimensional database. A set of SUM queries can be recorded as a system of linear equations of the form

$$MX = V, \tag{1}$$

where $X = (x_1, \ldots, x_n)$ and V is the vector with the values returned by the SUM queries corresponding to the rows of the matrix M. Each query corresponds to a row of the matrix M. It is enough to store only linearly independent queries in

the matrix M, since if several queries are known, then all their linear combinations are known too. The standard elementary row and column operations used to simplify systems of linear equations (1) result in a new system with rows corresponding to new queries with the outcomes equal to the corresponding values in the column V again. Therefore, we can assume that (1) has been simplified and stores a so-called *normalized query basis matrix*, so that $M = M_k$, where

$$M_k = (I_k | M'_k) \tag{2}$$

and I_k is a $(k \times k)$ identity matrix. Then the matrix M is said to be in a *normalized* form. The row vectors of M_k form a basis of the space of all queries with outcomes which are known, since they all can be derived using linear combinations of query vectors.

Audit Expert is a system using a normalized basis matrix to store all queries answered so far (cf. [15]). When a new query is added, Audit Expert adds it to the matrix and then reduces it to a normalized basis form again.

Theorem 1 ([15]). *The time complexity of the combinatorial algorithm dynamically processing the query matrix of the Audit Expert and maintaining it in a normalized form for a set of k consecutive queries is $O(k^2)$. The statistical database is compromised if and only if the normalized query basis matrix M_k has a row with exactly one nonzero entry.*

The paper [38] suggested using a Static Audit Expert, where the query basis matrix is fixed by the system (possibly the database administrator) in advance. A user's query is then allowed to be answered if it belongs to the vector space spanned by the rows of the matrix.

Theorem 2 ([38]). *The time complexity of the combinatorial algorithm for processing each new query by a Static Audit Expert with a predesigned query matrix in a normalized basis form is $O(k)$.*

This shows that Static Audit Expert is substantially more efficient than the dynamic Audit Expert. The maximum number of answerable queries, for databases where all SUM queries are posable, was determined in [35], where a combinatorial algorithm for constructing these sets of queries was also given.

Theorem 3 ([35]). (i) *In a 1-dimensional database with n real entries, the maximum number of SUM queries answerable without a compromise is equal to $\binom{n}{\lfloor n/2 \rfloor}$.*
(ii) ([24]) *The maximum is achieved if and only if the set of all entries is partitioned into two parts of size $\lfloor n/2 \rfloor$ and $\lceil n/2 \rceil$, and each allowed query set has equal numbers of elements from both parts.*

The *usability* of a statistical database is defined as the ratio of the maximum number of valid statistics that can be disclosed without a database compromise to the total number of valid statistics in the database.

If a confidential statistic based on one record has been revealed, then the term 1-compromise is used. The problem of preventing a compromise (1-compromise) can also be called the problem of preserving anonymity (1-anonymity).

Theorem 4 ([8, 23]). *In a statistical database of size n where all statistics are valid, the usability for 1-compromise is equal to $\binom{n}{\lfloor n/2 \rfloor}$.*

Theorem 5 ([11]). *In a statistical database where all statistics are valid and a fixed set of statistics are confidential and should not be disclosed, it is an NP-complete problem to decide whether the usability $\binom{n}{\lfloor n/2 \rfloor}$ can be achieved.*

Theorem 6 ([10]). *There exists a polynomial time algorithm to decide whether the usability $\binom{n}{\lfloor n/2 \rfloor}$ can be achieved in a statistical database where each statistic is based on at most two records, or each record appears in at most two statistics. It is an NP-complete problem to answer this question for statistical databases, where each statistic is based on exactly four records or each record appears in at most three statistics.*

Range queries are a special case of SUM queries. A *range query* is a sum of the form $a_1x_1 + \cdots + a_nx_n$, where $x_1, \ldots, x_n$ are values of the quantitative attribute A_0 in all records of the query set, and the query set is not arbitrary, but is selected using a range defined by inequalities as follows. Let $b_1, \ldots, b_m$ and $c_1, \ldots, c_m$ be real numbers such that $b_1 \leq c_1, \ldots, b_m \leq c_m$. A *query set* of a range query is a set of all records $(r_0, r_1, \ldots, r_m)$ of the database such that the following inequalities hold: $b_1 \leq r_1 \leq c_1, \ldots, b_m \leq r_m \leq c_m$. The value of the range query is the sum of the values of the quantitative attribute A_0 in all records in the query set.

The paper [9] presents several new results concerning the usability of statistical databases for general SUM, COUNT and MEAN queries as well as for the corresponding range queries, and combinatorial algorithms for constructing such sets of queries. In certain special cases the authors derive the usability of m-dimensional statistical databases for all $m \geq 1$.

The paper [3] is devoted to special sets of queries, where each record in a database is contained in at most two queries. Sets of queries of this sort are called *queries of type α*. For a set Q of queries, a graph $G = G(Q)$ is associated with Q. The vertices of $G(Q)$ correspond to the queries in Q and edges of $G(Q)$ correspond to records of the database. The authors of [3] introduce the notion of the *L-core* of the graph G. This concept makes it possible to formulate necessary and sufficient conditions for the set Q to be compromise free. The paper [3] shows how to determine the L-core from the eigenvalues of the graph, and proposes an algorithm for computing the L-core directly from the graph.

Theorem 7 ([3]). *Let Q be a query set of type α for a statistical database and let $G = G(Q)$ be the graph associated with Q. Then Q is compromise-free if and only if G coincides with its L-core.*

Several articles investigated range queries, where the values of the quantitative attribute A_0 are confidential and should not be compromised. For $i = 1, \ldots, k$, denote by d_i the number of distinct values of the characteristic attribute A_i in the database. The main result of the paper [27] shows that the largest set of all range queries, which does not lead to a compromise, is uniquely

determined and coincides with the set of all range queries with an even number of records.

Theorem 8 ([27]). *Let D be a k-dimensional database of size $d_1 \times \cdots \times d_k$. Then the usability of D is equal to $1-\frac{1}{2^k}\prod_{i=1}^{k} f(d_i)$, where $f(x) = (x+2)/(x+1)$ for x even, $f(x) = (x+1)/x$ for x odd.*

It follows that the usability of the database always belongs to the segment

$$\left[1 - \frac{1}{2^k}\prod_{i=1}^{k}\frac{d_i+2}{d_i+1}, 1 - \frac{1}{2^k}\prod_{i=1}^{k}\frac{d_i+1}{d_i}\right]. \tag{3}$$

In [5], a formula is given for the usability of range queries in a 1-dimensional database that is allowed to contain many indistinguishable copies of some records.

3 Relative compromise

A new type of compromise, which does not involve the disclosure of exact values, was introduced in [37]. Namely, a set S of records in a statistical database is said to be *relatively compromised* with respect to a field F if the relative order of magnitude of the F-values of the records in S becomes known [37]. It is shown in [37] that even when the exact confidential information remains protected, relative compromise may still be possible. Possible consequences of relative compromise are studied too. By applying block designs for the design of queries, it is shown in [37] that a relative compromise can be achieved even if the overlap of any two query sets is restricted not to exceed one element. In the case of SUM queries of fixed query size, the paper [37] used block designs to derive a number of conditions for the relative compromise to occur.

Theorem 9 ([38]). *Let D be a 1-dimensional database with n records, where SUM queries are allowed, and let M_k be the normalized query basis matrix of the Audit Expert. Then there is a relative compromise if and only if at least one of the following conditions is satisfied.*

(i) *There exists a row of the normalized query basis matrix containing exactly one nonzero element.*
(ii) *A row of the normalized query basis matrix contains exactly two nonzero entries which sum to zero.*
(iii) *There exist two rows $i \neq j$ of the normalized query basis matrix $M_k = (I_k|M'_k)$ such that the rows M'_i and M'_j are identical.*

The paper [38] classifies various types of compromise and extends the mechanism of Audit Expert to exclude relative compromises for SUM queries. The paper [36] used Audit Expert to determine the maximal number of answerable SUM queries preventing a relative compromise.

Theorem 10 ([36]). *Let D be a 1-dimensional database with n records, where SUM queries are allowed. The maximum number of SUM queries preventing relative compromised can be achieved by using a Static Audit Expert with the normalized query basis matrix $M_{n-1} = (I_{n-1}|M'_{n-1})$, where the transpose of M'_{n-1} is equal to*

$$(-[n/2], \ldots, -3, -2, 1, 2, 3, \ldots, [(n+1)/2]).$$

It follows from the results of [47] that the exact value of the maximum number of SUM queries in Theorem 10 coincides with the middle coefficient of the polynomial

$$(1+x)(1+x^2)\cdots(1+x^{[n/2]})(1+x)(1+x^2)\cdots(1+x^{[(n+1)/2]})$$

if the order of the polynomial is even, and it coincides with both of the two middle coefficients if the order of this polynomial is odd.

4 Group compromise or k-compromise

Statistics revealing information about a subset of k or fewer individuals may also need to be protected, because supplementary information often allows an attacker to derive private data about an individual from such statistics. The disclosure of a statistic based on k or fewer records in the database is called a *k-compromise*. The prevention of a k-compromise can also be called the preservation of k-anonymity.

It was shown in [24] that the usability of k-compromise in a statistical database with n records is in $O(n^{-1-k/2})$. Denote by $G(n,k)$ the maximum number of SUM queries, which can prevent k-compromise in the 1-dimensional database with n records. For any positive integers n and k, the next theorem determines the value $G(n,k)$ up to a constant factor less than $1/2$.

Theorem 11 ([1]). *If $n/2 \le k < n$, then*

$$\frac{k+1}{n}\binom{n}{k+1} < G(n,k) \le \binom{n}{k+1}.$$

If $2 \le k < n/2$ and n is odd, then

$$\frac{n+1}{n-1}\binom{n-1}{\frac{n-3}{2}} \le G(n,k) < 2\binom{n-1}{\frac{n-3}{2}}.$$

If $2 \le k < n/2$ and n is even, then

$$\frac{n+2}{2n-2}\binom{n}{\frac{n-2}{2}} \le G(n,k) < \binom{n}{\frac{n-2}{2}}.$$

Further, denote by $G(n,m,k)$ (resp., $G(n,\le m,k)$) the maximum number of SUM queries in the database, where each of the sums contains m (resp., at most m) summands, and k-compromise is prevented.

Theorem 12 ([16]). *Let $m \leq n$ be positive integers, and let $t = \lfloor n/m \rfloor$. Then the following conditions are satisfied.*

(i) *If $m \ll n$, then*

$$G(n, m, 1) = t\binom{n-t}{m-1}.$$

(ii) *If $n \to \infty$, then*

$$G(n, \leq m, 1) = t\binom{n-t}{m-1}(1 + o(1)).$$

Theorem 13 ([1]). *Let $k < m \leq n$ be positive integers, and let*

$$t \in \{\lfloor n/m \rfloor, \lfloor (n+1)/m \rfloor\}.$$

Then the following equality holds:

$$G(n, m, k) = t\binom{n-t}{m-1}.$$

Furthermore, the optimal set of SUM queries involving k summands corresponds to the set of $(0, 1)$-solutions of weight m to the linear equation

$$(m-1)x_1 + \cdots + (m-1)x_t - x_{t+1} - \cdots - x_n = 0. \tag{4}$$

If $\lfloor n/m \rfloor = \lfloor (n+1)/m \rfloor$, then the optimal set of SUM queries is unique up to permutation of the elements. If, however, $\lfloor n/m \rfloor \neq \lfloor (n+1)/m \rfloor$, then there exist precisely two (up to permutation of the elements) optimal sets of SUM queries determined by the linear equation (4) corresponding to $t = \lfloor n/m \rfloor$ or $t = \lfloor (n+1)/m \rfloor$, respectively.

Corollary 1 ([1]). *Let $k < m$ be positive integers. If $n \to \infty$, then*

$$G(n, \leq m, k) = t\binom{n-t}{m-1}(1 + o(1)).$$

Theorem 14 ([1]). *Let k, m and n be positive integers satisfying one of the following conditions:*

(i) *$m \geq 8$, $n \geq m^2$,*
(ii) *$4 \leq m \leq 7$ and $n \geq cm^2$ for a positive constant c,*

and let $t = \lfloor n/m \rfloor$. Then the following equality holds:

$$G(n, \leq m, k) = t\binom{n-t}{m-1}.$$

Conjecture 1 ([1]). If $n \geq 3$, then

$$G(n, 2) = \sum_{i=0}^{t} \binom{t}{i}\binom{n-t}{2i},$$

where $t = \lfloor n/3 \rfloor$.

In [12] a new result is obtained for the problem of maximizing the number of disclosed range queries preventing k-compromise, where k is odd, in the case of a 1-dimensional statistical database.

Theorem 15 ([12]). *Let D be a 1-dimensional database with n records, where range queries are allowed, and let $k = 2\ell - 1$, where $\ell > 1$ is a positive integer. Then the maximum number of elements in a k-compromise free set of range queries in D is equal to $\lfloor n/2\ell \rfloor (\lfloor n/\ell \rfloor - \lfloor n/2\ell \rfloor)$.*

The paper [13] determines the maximum number of sum totals that can be disclosed without leading to a 2-compromise in a 1-dimensional database for range queries. The following theorem was proved in [13] for a 1-dimensional statistical database of size n, where n is odd, or n is even and is greater than 52. For all other values of n, these formulas were proved in [32].

Theorem 16 ([13, 32]). *Let D be a 1-dimensional database with n records, where range queries are allowed. Then the maximum number $\mu_2(n)$ of elements in a 2-compromise free set of range queries in D is equal to*

$$\mu_2(n) = \begin{cases} (n+1)^2/16 & \text{for odd } n \geq 1; \\ n^2/16 & \text{for even } n \neq 12. \end{cases} \tag{5}$$

The paper [45] uses graphs to represent trust levels in informational relations among entities for the purposes of treating the requirements of access to confidential data for maintaining privacy and security.

The paper [44] introduces a Hippocratic security method for managing a collection of statistical databases by a virtual community at several institutions following a collection of management rules.

An attacker can often gain insight into confidential records stored in a statistical database using additional available information about the types of attributes stored in the database (called *working knowledge*), general restrictions on the values of the attributes in the real world (called *supplementary knowledge*), or additional restrictions on the values of attributes caused by various legal systems (called *legal knowledge*). The paper [39] proposed to use knowledge based systems capturing working knowledge, supplemental knowledge and legal knowledge to regulate access to statistical databases for the prevention of compromise.

5 Generalizations and other related results

A new method for maintaining the integrity of data in publicly accessible databases is developed in [26]. The method is based on the recent development of pseudo-random function families and sibling intractable function families.

A practical method for maintaining anonymous and verifiable databases with public data held in separate databases is introduced in [25]. It prevents unauthorized users from collecting and collating private data concerning individuals from these separate databases. The method is based on the use of smartcards

and the improved Leighton-Micali protocol for the distribution of keys and can be extended to mobile computing environments.

The security problems and possible mechanisms for the prevention of compromises are discussed in [34] with particular attention devoted to medical databases, where confidentiality is paramount. The paper concludes with a proposal for a security subsystem to be incorporated in a database management system. Applications of value added networks in managing the security of information stored in statistical databases in the health informatics sector are discussed in [40, 41].

It is explained in [20] that the multidimensional matrix model of statistical databases and the multidimensional cubes of On-Line Analytical Processing (OLAP) are essentially the same. The paper investigates the application of decision trees to mining information from statistical databases and studies robust noise addition methods to ensure the preservation of privacy. Methods for preserving privacy and enabling k-means clustering are proposed in [31, 43].

A novel noise addition framework for a statistical database containing several numerical attributes and a single categorical attribute is studied in [28]. Data perturbation techniques for the prevention of disclosure of confidential values are studied in [29] in order to handle categorical attributes without a natural order of their values. A novel approach towards clustering of such categorical values is proposed in [29] and is used to perturb data. It applies horizontal partitioning and clusters the values of a given categorical attribute rather than the records of the datasets. An experimental study was performed to compare the resulting perturbation system DETECTIVE in its effectiveness with another system called CACTUS [29].

Notice that k-anonymity is a broad concept applicable in various settings. For example, in [14] it is studied for recommendation systems. It is shown in [19] that permutation is the most essential principle underlying any anonymization of microdata that involves the utility and privacy guarantees. Any anonymization for microdata can be regarded as a permutation combined with the possible addition of a small amount of noise. This lead to a new natural privacy model called (d, v, f)-permuted privacy. It incorporates subject-verifiability, i.e., the ability of every subject supplying original data to verify privacy.

The paper [46] explains how sum labellings of graphs can be used for representing the access structure of a secret sharing scheme. Another privacy-preserving framework using novel noise addition techniques is investigated in [30]. It uses noise addition to categorical values as well, so that attributes of all types are protected. An experimental study of the practical system VICUS incorporating noise addition for categorical attributes is carried out in [22].

Statistical disclosure control is also discussed in [21], where a strategic dependency model of a statistical data warehouse system is proposed and an associated model of trust is explored.

The problem of evaluating and comparing privacy provided by various techniques is tackled in [2], where a novel entropy based security measure is proposed. It can be applied to any generalization, restriction or data modification

technique for preserving privacy of statistical databases. This measure is used in [2] in an empirical study evaluating and comparing the methods of query restriction, sampling and noise addition.

A new method for achieving k-anonymity of network graph data prior to its release is considered in [7] for privacy protection. The method is based on randomizing the location of the triangles in the graph. It is shown that this new method preserves the main structural characteristics of the graph, which can provide valuable information for the study of the graph, while preserving k-anonymity.

Acknowledgements

The authors are grateful to three reviewers for comments and corrections that have helped to improve this paper. This work has been supported by Discovery grant DP160100913 from Australian Research Council.

References

1. Ahlswede, R., Aydinian, H.: On security of statistical databases. SIAM Journal on Discrete Mathematics 25, 1778–1791 (2011)
2. Alfalayleh, M., Brankovic, L.: Quantifying privacy: A novel entropy-based measure of disclosure risk. In: 25th International Workshop on Combinatorial Algorithms, IWOCA 2014. Lecture Notes in Computer Science, vol. 8986, pp. 24–36 (2015)
3. Brankovic, L., Cvetković, D.: The eigenspace of the eigenvalue −2 in generalized line graphs and a problem in security of statistical databases. Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat. 14, 37–48 (2003)
4. Brankovic, L., Giggins, H.: Statistical database security. In: Security, Privacy, and Trust in Modern Data Management, pp. 167–181. Data-Centric Systems and Applications, Springer (2007)
5. Brankovic, L., Horak, P., Miller, M.: An optimization problem in statistical database security. SIAM Journal on Discrete Mathematics 13(3), 346–353 (2000)
6. Brankovic, L., Islam, M.Z., Giggins, H.: Privacy-preserving data mining. In: Security, Privacy, and Trust in Modern Data Management, pp. 151–165. Data-Centric Systems and Applications, Springer (2007)
7. Brankovic, L., Lopez, M., Miller, M., Sebe, F.: Triangle randomization for social network data anonymization. Ars Mathematica Contemporanea 7, 461–477 (2014)
8. Brankovic, L., Miller, M.: An application of combinatorics to the security of statistical databases. Austral. Math. Soc. Gaz. 22(4), 173–177 (1995)
9. Brankovic, L., Miller, M., Horak, P., Wrightson, G.: Usability of compromise-free statistical databases. In: Proceedings of the International Working Conference on Scientific and Statistical Database Management. pp. 144–154. (Melbourne, Australia, January 29-30) (1997)
10. Brankovic, L., Miller, M., Širáň, J.: Graphs, 0-1 matrices, and usability of statistical databases. Congr. Numer. 120, 169–182 (1996)
11. Brankovic, L., Miller, M., Širáň, J.: Towards a practical auditing method for the prevention of statistical database compromise. In: Proceedings of the Seventh Australasian Database Conference. pp. 177–184. (Melbourne, Australia, January 29-30) (1996)

12. Brankovic, L., Miller, M., Širáň, J.: On range query dsability of statistical databases. Int. J. Comput. Math. 79(12), 1265–1271 (2002)
13. Brankovic, L., Širáň, J.: 2-compromise usability in 1-dimensional statistical databases. In: Computing and Combinatorics, Lecture Notes in Comput. Sci., vol. 2387, pp. 448–455. Springer, Berlin (2002)
14. Casino, F., Domingo-Ferrer, J., Patsakis, C., Puig, D., Solanas, A.: A k-anonymous approach to privacy preserving collaborative filtering. Journal of Computer and System Sciences 81, 1000–1011 (2015)
15. Chin, F.Y., Ozsoyoglu, G.: Auditing and inference control in statistical databases. IEEE Transactions on Software Engineering 8.6, 574–582 (1982)
16. Demetrovics, J., Katona, G.O.H., Miklos, D.: On the security of individual data. In: Foundations of Information and Knowledge Systems. Lecture Notes in Computer Science, vol. 2942, pp. 49–58 (2004)
17. Domingo-Ferrer, J.: Inference Control in Statistical Databases. Springer, Berlin, 6 edn. (2002)
18. Domingo-Ferrer, J.: A survey of inference control methods for privacy-preserving data mining. In: Privacy-Preserving Data Mining Models and Algorithms, Advances in Database Systems, vol. 34, pp. 53–80. Springer (2008)
19. Domingo-Ferrer, J., Muralidhar, K.: New directions in anonymization: Permutation paradigm, verifiability by subjects and intruders, transparency to users. Information Sciences 337-338, 11–24 (2016)
20. Estivill-Castro, V., Brankovic, L.: Data swapping: Balancing privacy against precision in mining for logic rules. In: Data Warehousing and Knowledge Discovery. Lecture Notes in Computer Science, vol. 1676, pp. 389–398 (1999)
21. Giggins, H., Brankovic, L.: Statistical disclosure control: To trust or not to trust. In: Proceedings of the International Symposium on Computer Science and its Applications. pp. 108–113. IEEE Computer Sociaty (2008)
22. Giggins, H., Brankovic, L.: VICUS – a noise addition technique for categorical data. In: Proceedings of the Tenth Australasian Data Mining Conference, AusDM 2012. Conferences in Research and Practice in Information Technology (CRPIT), vol. 134, pp. 139–148 (2012)
23. Griggs, J.R.: Concentrating subset sums at k points. Bull. Inst. Combin. Appl. 20, 65–74 (1997)
24. Griggs, J.R.: Database security and the distribution of subset sums in R^m. In: Graph Theory and Combinatorial Biology. Bolyai Soc. Math. Stud., vol. 7, pp. 223–252 (1997)
25. Hardjono, T., Seberry, J.: Applications of smartcards for anonymous and verifiable databases. Computers and Security 14, 465–472 (1995)
26. Hardjono, T., Zheng, Y., Seberry, J.: Database authentication revisited. Computers and Security 13, 573–580 (1994)
27. Horak, P., Brankovic, L., Miller, M.: A combinatorial problem in database security. Discrete Appl. Math. 91(1-3), 119–126 (1999)
28. Islam, M.Z., Brankovic, L.: A framework for privacy preserving classification in data mining. In: Proceedings of the 2nd Workshop on Australasian Information Security. Data Mining and Web Intelligence, and Software Internationalisation, vol. 32, pp. 163–168 (2004)
29. Islam, M.Z., Brankovic, L.: DETECTIVE: a decision tree based categorical value clustering and perturbation technique for preserving privacy in data mining. In: Proceedings of the 3rd IEEE International Conference on Industrial Informatics, INDIN 2005. pp. 701–708 (2005)

30. Islam, M.Z., Brankovic, L.: Privacy preserving data mining: A noise addition framework using a novel clustering technique. Knowledge-Based Systems 24, 1214–1223 (2011)
31. Liu, D., Bertino, E., Yi, X.: Privacy of outsourced k-means clustering. In: Proceedings of the 9th ACM Symposium on Information, Computer and Communications Security, ASIA CCS 2014. pp. 123–133 (2014)
32. Mathieson, L., King, T., Brankovic, L.: 2-compromise: usability in 1-dimensional statistical database. Research Gate, avalilable online at https://www.researchgate.net/publication/228973056 (2008)
33. Miller, M.: A model of statistical databse compromise incorporating supplementary knowledge. In: Databases in the 1990's. pp. 258–267 (1991)
34. Miller, M., Cooper, J.: Security considerations for present and future medical databases. International Journal of Medical Informatics 41, 39–46 (1996)
35. Miller, M., Roberts, I., Simpson, J.: Application of symmetric chains to an optimization problem in the security of statistical databases. Bull. Inst. Combin. Appl. 2, 47–58 (1991)
36. Miller, M., Roberts, I., Simpson, J.: Prevention of relative compromise in statistical databases using audit expert. Bull. Inst. Combin. Appl. 10, 51–62 (1994)
37. Miller, M., Seberry, J.: Relative compromise of statistical databases. The Australian Computer Journal 21(2), 56–61 (1989)
38. Miller, M., Seberry, J.: Audit expert and statistical database security. In: Databases in the 1990's. pp. 149–174 (1991)
39. Mishra, V., Stranieri, A., Miller, M., Ryan, J.: Knowledge based regulation of statistical databases. WSEAS Transactions on Information Science and Applications 3(2), 239–244 (2006)
40. Pacheco, F., Cooper, J., Bomba, D., Morris, S., Miller, M., Brankovic, L.: Education issues in health informatics. Informatics in Healthcare 4, 101–105 (1995)
41. Pacheco, F., Cooper, J., Bomba, D., Morris, S., Miller, M., Brankovic, L.: Value added networks (VANs) and their benefit to a health information system. Informatics in Healthcare 4, 141–144 (1995)
42. Pieprzyk, J., Hardjono, T., Seberry, J.: Fundamentals of Computer Security. Springer-Verlag, Berlin (2003)
43. Rao, F.Y., Samanthula, B., Bertino, E., Yi, X., Liu, D.: Privacy-preserving and outsourced multi-user k-means clustering. In: Proceedings of the IEEE Conference on Collaboration and Internet Computing, CIC 2015. pp. 80–89 (2015)
44. Skinner, G., Chang, E., McMahon, M., Aisbett, J., Miller, M.: Shield privacy hippocratic security method for virtual community. In: IECON Proceedings (Industrial Electronics Conference). pp. 472–479 (2004)
45. Skinner, G., Miller, M.: Managing privacy, trust, security, and context relationships using weighted graph representations. WSEAS Transactions on Information Science and Applications 3(2), 283–290 (2006)
46. Slamet, S., Sugeng, K.A., Miller, M.: Sum graph based access structure in a secret sharing scheme. Journal of Prime Research in Mathematics 2, 113–119 (2006)
47. Stanley, R.P.: Weyl groups, the hard Lefshetz theorem, and the Sperner property. SIAM J. Alg. Disc. Meth. 1, 168–184 (1980)
48. Yi, X., Paulet, R., Bertino, E.: Private Information Retrieval. Morgan and Claypool, United States (2013)